PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00

A2

(11) International Publication Number: **WO 00/58473**

(43) International Publication Date: 5 October 2000 (05.10.00)

(21) International Application Number: PCT/US00/08621

(22) International Filing Date: 31 March 2000 (31.03.00)

(30) Priority Data:

60/127,607	31 March 1999 (31.03.99)	US
60/127,636	2 April 1999 (02.04.99)	US
60/127,728	5 April 1999 (05.04.99)	US
09/540,763	30 March 2000 (30.03.00)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/127,607 (CIP)
Filed on	31 March 1999 (31.03.99)
US	60/127,636 (CIP)
Filed on	2 April 1999 (02.04.99)
US	60/127,728 (CIP)
Filed on	5 April 1999 (05.04.99)
US	09/540,763 (CIP)
Filed on	30 March 2000 (30.03.00)

(71) Applicant *(for all designated States except US)*: CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven, CT 06516 (US). LEACH, Martin [GB/US]; 884 School Street, Webster, MA 01570 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"

(57) Abstract

The present invention provides open reading frames ORFX, encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

2190	87639197 (4379, 4380)	Novel Protein sim. GBank gi|132575|sp|P29315|RINI_RAT - RIBONUCLEASE INHIBITOR		nucleaseinhib	22278996, 22278999, 29331822, 29331824, 29331826, 265008, 264910, 60170831, 55812038, 52644296, 265010, 265018, 264685, 264688, 56181562, 21906769, 35695917, 265022, 60170394, 22279000
2191	95198928 (4381, 4382)	Novel Protein sim. GBank gi|5327002|emb|CAB46272.1| - (Y18503) XAP-5-like protein [Homo sapiens]			29331825, 29331826, 29331830, 264510, 264511, 264910, 264593, 264594, 264556, 264559
2192	11126316 (4383, 4384)	Novel Protein sim. GBank gi|462600|sp|P34400|MI10_CAEEL - MIG-10 PROTEIN	Contains protein domain (PF00169) - PH domain		264558
2193	94140073 (4385, 4386)	Novel Protein sim. GBank gi|5420389|emb|CAB46680.1| - (AJ243480) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	56181686, 29331825, 29331827, 264508, 264909, 265008, 264592, 60432229, 264288, 264684, 264766, 35695917, 33657023, 60431602, 60431528, 55810764, 55811576, 65274791, 35695855, 60431850, 56182323, 60432113
2194	21418714 (4387, 4388)	Novel Protein sim. GBank gi|2773341 (AF040954) - putative protein phosphatase 1 nuclear targeting subunit [Rattus norvegicus]			264592
2195	88083023 (4389, 4390)	Novel Protein sim. GBank gi|2832763|emb|CAA15685.1| - (AL009191) /prediction=(method:; /prediction=(method:; /match=(desc:; /match=(desc:; /motif=(desc: [Drosophila melanogaster]		UNCLASSIFIED	22278996, 22278999, 35696052, 265006, 21906754, 265017, 35695917, 265021, 265022, 35695855
2196	85091631 (4391, 4392)	Novel Protein sim. GBank gi|5262487|emb|CAB45699.1| - (AL080076) hypothetical protein [Homo sapiens]		collagen	56182575, 35696286, 22278997, 22278999, 264259, 29331822, 66714117, 60432289, 29331827, 35696052, 29331828, 264508, 52644045, 56182435, 264510, 265007, 265008, 265009, 60433438, 55812038, 265010, 265011, 264448, 264288, 264686, 264687, 52644229, 21906765, 21906766, 21906767, 35695917, 265022, 264691, 33657023, 264693, 18108370, 18108376, 35696423, 55811576, 65274791, 35695855, 264636, 56182323, 18108385
2197	95073813 (4393, 4394)	Novel Protein sim. GBank gi|4929567|gb|AAD34044.1|AF15180 - (AF151807) CGI-49 protein [Homo sapiens]			264768, 264769, 21906765, 21906766, 21906767, 29148627, 55811957, 35696286, 265020, 22278998, 265021, 264259, 33657023, 264693, 29331824, 35696052, 29331828, 18108370, 35695855, 264113, 265008, 264910, 60432229, 56182323, 33657402, 264758, 83373044, 21906754, 265018, 265019, 22279002, 264482, 264448, 264565, 264288, 264369
2198	88060914 (4395, 4396)	Novel Protein sim. GBank gi|3548787 (AC005622) - R30953_1 [Homo sapiens]		UNCLASSIFIED	

1015	95418879 (2029, 2030)	Novel Protein sim. GBank gi|4159995 (AF063095) - SELIL [Mus musculus]	Contains protein domain (PF00040) - Fibronectin type II domain	struct	22278994, 22278995, 56994075, 22278996, 22278999, 264259, 29331825, 29331828, 264907, 56182435, 264510, 264591, 264593, 60433356, 264594, 55812038, 264758, 21906754, 33657084, 265010, 264600, 265017, 265018, 265019, 18108351, 21906765, 21906766, 21906767, 21906768, 55811957, 265022, 33657023, 65274620, 33657182, 32833986, 18108370, 18108377, 55811576, 35696423, 264630, 22279000, 264565
1016	79559694 (2031, 2032)	Novel Protein sim. GBank gi|2506969|sp|P41407|ACPD_ECOLI - ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)		esterase	264686, 264693
1017	11069213 (2033, 2034)	Novel Protein sim. GBank gi|5103943|dbj|BAA79259.1| - (AP000059) 802aa long hypothetical oligopeptide-binding protein oppA [Aeropyrum pernix]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	transport	264600
1018	80072430 (2035, 2036)	Novel Protein sim. GBank gi|4493973|emb|CAB39032.1| - (AL034559) predicted using hexExon; MAL3P7.14 (PFC0925w), Hypothetical protein, len: 469 aa [Plasmodium falciparum]			22278996, 29148627, 264563
1019	11703807 (2037, 2038)			UNCLASSIFIED	264686
1020	80234432 (2039, 2040)				264508, 264509, 264512, 264600, 264762, 264769, 264689, 18108370, 264636, 264638, 264486
1021	37036243 (2041, 2042)	Novel Protein sim. GBank gi|4633807|gb|AAD26859.1|AF12779 - (AF127795) trehalose biosynthetic enzyme TreY [Rhizobium leguminosarum bv. viciae]		synthase	264769
1022	80502627 (2043, 2044)	Novel Protein sim. GBank gi|1781230|emb|CAB06277| - (Z83867) hypothetical protein Rv3137 [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	phosphatase	35696052, 264508, 265008, 265009, 264769, 18108387, 264563
1023	11399341 (2045, 2046)	Novel Protein sim. GBank gi|3777495 (U92083) - calcium transporting ATPase [Pichia angusta]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase_associated	264593
1024	80057129 (2047, 2048)			UNCLASSIFIED	52646842, 33657402, 33657023, 18108379, 55811576, 264631, 264556, 264557, 264559, 18108388, 264566
1025	79644200 (2049, 2050)	Novel Protein sim. GBank gi|3483045|emb|CAA20556| - (AL031371) putative transport system permease protein [Streptomyces coelicolor]		transport	264693
1026	80025946 (2051, 2052)	Novel Protein sim. GBank gi|1174922|sp|Q02322|UVRD_HAEIN - DNA HELICASE II		helicase	264602
1027	17659234 (2053, 2054)	Novel Protein sim. GBank gi|4757728|ref|NP_004886.1|pAGTA - angiotensin/vasopressin receptor AII/AVP-like		UNCLASSIFIED	265017

8. The vector of claim 7, wherein said vector is an expression vector.

9 A host cell comprising the isolated nucleic acid molecule of claim 1.

10. A substantially purified polypeptide comprising an amino acid sequence at least 80% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2*n*, wherein *n* is any integer 1-3161.

11. The polypeptide of claim 10, wherein said polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 2*n*, wherein *n* is any integer 1-3161.

12. An antibody that selectively binds to the polypeptide of claim 10.

13. A pharmaceutical composition comprising a therapeutically or prophylactically effective amount of a therapeutic selected from the group consisting of:

a) the nucleic acid of claim 1;

b) the polypeptide of claim 10; and

c) the antibody of claim 12;

and a pharmaceutically acceptable carrier.

14. A kit comprising in one or more containers, a therapeutically or prophylactically effective amount of the pharmaceutical composition of claim 13.

15. A method of producing the polypeptide of claim 10, said method comprising culturing the host cell of claim 9 under conditions in which the nucleic acid molecule is expressed.

16. A method of detecting the presence of the polypeptide of claim 10 in a sample, comprising contacting the sample with a compound that selectively binds to said polypeptide under conditions allowing the formation of a complex between said polypeptide and said

What is claimed is:

1. An isolated nucleic acid molecule encoding a polypeptide comprising an amin acid sequence that is at least 85% identical to a polypeptide including an amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is any integer 1-3161, or the complement thereof.

2. The isolated nucleic acid molecule of claim 1, said molecule hybridizing under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule comprising the sequence of nucleotides selected from the group consisting of SEQ ID NO:2*n*- wherein *n* is any integer 1-3161, or the complement thereof.

3. The isolated nucleic acid molecule of claim 1, said molecule encoding a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ I NO: 2*n*, wherein n is any integer 1-3161, or an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of S ID NO: 2*n*.

4. The isolated nucleic acid molecule of claim 1, wherein said molecule encodes polypeptide comprising the amino acid sequence selected from the group consisting of SEQ I NO: 2*n*, wherein *n* is any integer 1-3161.

5. The isolated nucleic acid molecule of claim 1, wherein said molecule comprise the sequence of nucleotides selected from the group consisting of SEQ ID NO:2*n*-1, wherein any integer 1-3161, or the complement thereof.

6. An oligonucleotide less than 100 nucleotides in length and comprising at least contiguous nucleotides selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is a integer 1-3161, or the complement thereof.

7. A vector comprising the nucleic acid molecule of claim 1.

compound, and detecting said complex, if present, thereby identifying said polypeptide in said sample.

17. A method of detecting the presence of a nucleic acid molecule of claim 1 in a sample, the method comprising contacting the sample with a nucleic acid probe or primer that selectively binds to the nucleic acid molecule and determining whether the nucleic acid probe c primer bound to the nucleic acid molecule of claim 1 is present in the sample.

18. A method for modulating the activity of the polypeptide of claim 10, the methoc comprising contacting a cell sample comprising the polypeptide of claim 10 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptid

19. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a ORFX-associated disorder, wherein said therapeutic is selected fro the group consisting of:

a) the nucleic acid of claim 1;

b) the polypeptide of claim 10; and

c) the antibody of claim 12.

20. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) contacting a test compound with the polypeptide of claim 10; and

b) determining if said test compound binds to said polypeptide,

wherein binding of said test compound to said polypeptide indicates the test compound is a modulator of activity or of latency or predisposition to an ORFX-associated disorder.

21. A method for screening for a modulator of activity or of latency or predispositio to an ORFX-associated disorder, said method comprising:

a) administering a test compound to a test subject at an increased risk ORFX-associated disorder, wherein said test subject recombinantly expresses a polypeptide encoded by the nucleotide of claim 1;

b) measuring expression the activity of said protein in said test subject;

c) measuring the activity of said protein in a control subject that recombinantly expresses said protein and is not at increased risk for an ORFX-associated disorder; and

d) comparing expression of said protein in said test subject and said control subject, wherein a change in the activity of said protein in said test subject relative to said control subject indicates the test compound is a modulator or of latency of predispostition to an ORFX-associated disorder.

22. The method of claim 20, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

23. A method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide of claim 11 in a subject, the method comprising:

a) measuring the amount of the polypeptide in a sample from said subject; and

b) comparing the amount of said polypeptide in step (a) to the amount of the polypeptide present in a control sample,

wherein an alteration in the level of the polypeptide in step (a) as compared to the control sample indicates the presence of or predisposition to a disease in said subject.

24. The method of claim 23, wherein said subject is a human.

25. A method for determining the presence of or predisposition to a disease associated with altered levels the nucleic acid molecule of claim 1 in a subject, the method comprising:

a) measuring the amount of the nucleic acid in a sample from the mammalian subject; and

b) comparing the amount of said nucleic acid in step (a) to the amount of the nucleic acid present in a control sample,

wherein an alteration in the level of the nucleic acid in step (a) as compared to the co sample indicates the presence of or predisposition to said disease in said subject.

26. The method of claim 25, wherein said subject is a human.

27. A method of treating or preventing a pathological condition associated with a ORFX-associated disorder in a subject, the method comprising administering to said subject polypeptide of claim 10 in an amount sufficient to alleviate or prevent said pathological condition.

28. The method of claim 27, wherein said subject is a human.

29. A method of treating or preventing a pathological condition associated with a ORFX-associated disorder in a subject, the method comprising administering to said subject nucleic acid molecule of claim 1 in an amount sufficient to alleviate or prevent said patholog condition.

30. The method of claim 29, wherein said subject is a human.

31. A method of treating or preventing a pathological condition associated with a ORFX-associated disorder in a subject, the method comprising administering to said subject antibody of claim 12 in an amount sufficient to alleviate or prevent said pathological conditio

32. The method of claim 31, wherein said subject is a human.

<212> DNA
<213> Homo sapiens

<400> 2051
gagcaaaact atcgttctac cggcaatatt ctgaaaagtg ccaaccaact tatttcgaat
60
aatagtgatc gtctcggtaa gaatttatgg accgacggtg aaatggggga gccagtaggt
120
atttatgcag catttaatga attagatgag gcaaaatttg tggcgtctca aatccaaaat
180
tgggtagatg atggtgggga attagatgat tgtgctgttt tatatcgtag taatagccaa
240
tctcgtgtta ttgaagaagc cttgattcgt tgccaaattc cttatcgaat ttatggcggg
300
atgcgattct tcgaacgcca agaaattaaa gatgcgttgg catatttacg tttaattaat
360
aatcgtcaag atgatgccgc atttgagcgt gtgattaata cgcctacgcg t
411

<210> 2052
<211> 137
<212> PRT
<213> Homo sapiens

<400> 2052
Glu Gln Asn Tyr Arg Ser Thr Gly Asn Ile Leu Lys Ser Ala Asn Gln
 1   5   10   15
Leu Ile Ser Asn Asn Ser Asp Arg Leu Gly Lys Asn Leu Trp Thr Asp
            20   25   30
Gly Glu Met Gly Glu Pro Val Gly Ile Tyr Ala Ala Phe Asn Glu Leu
        35   40   45
Asp Glu Ala Lys Phe Val Ala Ser Gln Ile Gln Asn Trp Val Asp Asp
    50   55   60
Gly Gly Glu Leu Asp Asp Cys Ala Val Leu Tyr Arg Ser Asn Ser Gln
65   70   75   80
Ser Arg Val Ile Glu Glu Ala Leu Ile Arg Cys Gln Ile Pro Tyr Arg
                85   90   95
Ile Tyr Gly Gly Met Arg Phe Phe Glu Arg Gln Glu Ile Lys Asp Ala
            100   105   110
Leu Ala Tyr Leu Arg Leu Ile Asn Asn Arg Gln Asp Asp Ala Ala Phe
        115   120   125
Glu Arg Val Ile Asn Thr Pro Thr Arg
    130   135

<210> 2053
<211> 287
<212> DNA
<213> Homo sapiens

<400> 2053
nccatggaag ccttcaatct tgtaagagaa agtgaacagc tgttttccat atgccaaatc
60
ccgctcctct gctggatcct gtgtaccagt ctgaagcaag agatgcagaa aggaaaagac
120

ctggccctga cctgccagag cactacctct gtgtactcct ctttcgtctt taacctgttc
180
acacctgagg gtgccgaggg cccgactccg caaacccagc accagctgaa ggccctgtgc
240
tccctggctg cagagggtat gtggacagac acatttgagt tttgtga
287

<210> 2054
<211> 79
<212> PRT
<213> Homo sapiens

<400> 2054
Ile Cys Gln Ile Pro Leu Leu Cys Trp Ile Leu Cys Thr Ser Leu Lys
1 5 10 15
Gln Glu Met Gln Lys Gly Lys Asp Leu Ala Leu Thr Cys Gln Ser Thr
20 25 30
Thr Ser Val Tyr Ser Ser Phe Val Phe Asn Leu Phe Thr Pro Glu Gly
35 40 45
Ala Glu Gly Pro Thr Pro Gln Thr Gln His Gln Leu Lys Ala Leu Cys
50 55 60
Ser Leu Ala Ala Glu Gly Met Trp Thr Asp Thr Phe Glu Phe Cys
65 70 75

<210> 2055
<211> 298
<212> DNA
<213> Homo sapiens

<400> 2055
nnacgcgttg ttatgaacaa tgacggtgtc ctctaccccg atacctgcgt gggtactgat
60
tcccacacca ccatggaaaa tggtcttggc attctgggct ggggcgtcgg tggtattgaa
120
gccgaggctg ctatgcttgg ccagcccatc tccatgctta tcccccgtgt tgttggcttt
180
aaacttactg gccaaacaca gccgggtgtc accgctacag atgttgttct taccattact
240
gatatgcttc gccagcatgg tgtgggtgga aaattcgggg aattctatgg gggaagcg
298

<210> 2056
<211> 99
<212> PRT
<213> Homo sapiens

<400> 2056
Xaa Arg Val Val Met Asn Asn Asp Gly Val Leu Tyr Pro Asp Thr Cys
1 5 10 15
Val Gly Thr Asp Ser His Thr Thr Met Glu Asn Gly Leu Gly Ile Leu
20 25 30
Gly Trp Gly Val Gly Gly Ile Glu Ala Glu Ala Ala Met Leu Gly Gln
35 40 45
Pro Ile Ser Met Leu Ile Pro Arg Val Val Gly Phe Lys Leu Thr Gly

ctgctggcgg cgggcgaggc cggcaccttc gacgtggccg tggtggatgc ggacaaggag
480
aactgctccg cctactacga gcgctgcctg cagctgctgc gacccggagg catcctcgcc
540
gtcctcagag tcctgtggcg cgggaaggtg ctgcaacctc cgaaagggga cgtggcggcc
600
gagtgtgtgc gaaacctaaa cgaacgcatc cggcgggacg tcagggtcta catcagcctc
660
ctgcccctgg gcgatggact caccttggcc ttcaagatct agggctggcc cctagtgagt
720
gggctcgagg gagggttgcc tgggaacccc aggaattgac cctgagtttt aaattcgaaa
780
ataaagtggg gctgggacac acgaaaaaaa aa
812

<210> 4378
<211> 233
<212> PRT
<213> Homo sapiens

<400> 4378
Xaa Leu Gly Arg Arg Cys Pro Pro Trp Arg Gly Arg Arg Glu Gln Gly
1 5 10 15
Leu Leu Pro Pro Glu Asp Ser Arg Leu Trp Gln Tyr Leu Leu Ser Arg
20 25 30
Ser Met Arg Glu His Pro Ala Leu Arg Ser Leu Arg Leu Leu Thr Leu
35 40 45
Glu Gln Pro Gln Gly Asp Ser Met Met Thr Cys Glu Gln Ala Gln Leu
50 55 60
Leu Ala Asn Leu Ala Arg Leu Ile Gln Ala Lys Lys Ala Leu Asp Leu
65 70 75 80
Gly Thr Phe Thr Gly Tyr Ser Ala Leu Ala Leu Ala Leu Ala Leu Pro
85 90 95
Ala Asp Gly Arg Val Val Thr Cys Glu Val Asp Ala Gln Pro Pro Glu
100 105 110
Leu Gly Arg Pro Leu Trp Arg Gln Ala Glu Ala Glu His Lys Ile Arg
115 120 125
Leu Arg Leu Lys Pro Ala Leu Glu Thr Leu Asp Glu Leu Leu Ala Ala
130 135 140
Gly Glu Ala Gly Thr Phe Asp Val Ala Val Val Asp Ala Asp Lys Glu
145 150 155 160
Asn Cys Ser Ala Tyr Tyr Glu Arg Cys Leu Gln Leu Leu Arg Pro Gly
165 170 175
Gly Ile Leu Ala Val Leu Arg Val Leu Trp Arg Gly Lys Val Leu Gln
180 185 190
Pro Pro Lys Gly Asp Val Ala Ala Glu Cys Val Arg Asn Leu Asn Glu
195 200 205
Arg Ile Arg Arg Asp Val Arg Val Tyr Ile Ser Leu Leu Pro Leu Gly
210 215 220
Asp Gly Leu Thr Leu Ala Phe Lys Ile
225 230

<210> 4379
<211> 2347

<212> DNA
<213> Homo sapiens

<400> 4379
ngaggaccaa gccatgcgtg cctttgagct aatgaggagc aacgcggccc tgttccagct
60
gggctcggcc ccgcggtgtg ctggattgtg tgcacgactc tgaagctgca gatggagaag
120
ggggaggacc cggtccccac ctgcctcacc cgcacggggc tgttcctgcg tttcctctgc
180
agccggttcc cgcggggcgc acagctgcgg ggcgcgctgc ggacgctgag cctcctggcc
240
gcgcagggcc tgtgggcgca gacgtccgtg cttcaccgag aggatctgga aaggctcggg
300
gtgcaggagt ccgacctccg tctgttcctg gacggagaca tcctccgcca ggacagagtc
360
tccaaaggct gctactcctt catccacctc agcttccagc agtttctcac tgccctgttc
420
tacaccctgg agaaggagga ggaagaggat agggacggcc acacctggga cattggggac
480
gtacagaagc tgctttccgg agtagaaaga ctcaggaacc ccgacctgat ccaagcaggc
540
tactactcct ttggcctcgc taacgagaag agagccaagg agttggaggc cacttttggc
600
tgccggatgt caccggacat caaacaggaa ttgctgcgat gcgacataag ttgtaagggt
660
ggacattcaa cggtgacaga cctgcaggag ctgctcggct gtctgtacga gtctcaggag
720
gaggagctgg tgaaggaggt gatggctcag ttcaaagaaa tatcctgca cttaaatgca
780
gtagacgttg tgccatcttc attctgcgtc aagcactgtc gaaacctgca gaaaatgtca
840
ctgcaggtaa taaaggagaa tctcccggag aatgtcactg cgtctgaatc agacgccgag
900
gttgagagat cccaggatga tcagcacatg cttcctttct ggacggacct ttgttccata
960
tttggatcaa ataaggatct gatgggtcta gcaatcaatg atagctttct cagtgcctcc
1020
ctagtaagga tcctgtgtga acaaatagcc tctgacacct gtcatctcca gagagtggtg
1080
ttcaaaaaca tttccccagc tgatgctcat cggaacctcn tgcctnnagc tcttcgaggt
1140
cacaagactg taacgtatct gacccttcaa ggcaatgacc aggatgatat gtttcccgca
1200
ttgtgtgagg tcttgagaca tccagaatgt aacctgcgat atctcgggtt ggtgtcttgt
1260
tccgctacca ctcagcagtg ggctgatctc tccttggccc ttgaagtcaa ccagtccctg
1320
acgtgcgtaa acctctccga caatgagctt ctggatgagg gtgctaagtt gctgtacaca
1380
actttgagac accccaagtg ctttctgcag aggttgtcgt tggaaaactg tcaccttaca
1440
gaagccaatt gcaaggacct tgctgctgtg ttggttgtca gccgggagct gacacacctg
1500

tgcttggcca agaaccccat tgggaataca ggggtgaagt ttctgtgtga gggcttgagg
1560
taccccgagt gtaaactgca gaccttggtg ctttggaact gcgacataac tagcgatggc
1620
tgctgcgatc tcacaaagct tctccaagaa aaatcaagcc tgttgtgttt ggatctgggg
1680
ctgaatcaca taggagttaa gggaatgaag ttcctgtgtg aggctttgag gaaaccactg
1740
tgcaacttga gatgtctgtg gttgtgggga tgttccatcc ctccgttcag ttgtgaagac
1800
gtctgctctg ccctcagctg caaccagagc ctcgtcactc tggacctggg tcagaatccc
1860
ttggggtcta gtggagtgaa gatgctgttt gaaaccttga catgttccag tggcaccctc
1920
cggacactca ggttgaaaat agatgacttt aatgatgaac tcaataagct gctggaagaa
1980
atagaagaaa aaaacccaca actgattatt gatactgaga aacatcatcc ctgggaagaa
2040
aggccttctt ctcatgactt catgatctga atccccccga gtcattcatt ctccatgaag
2100
tcatcgattt tccaggtgtg ggtgaactgc ctgtgactcc tctcctcccc cgcccctacc
2160
cctcagggat aatgagttca ttgctgggct agatgtttta gccatgattc tgcctctgtt
2220
ttatacctgc acacatcctt atctttgtta catatgaaat atctgtatca cgggtatatt
2280
gagagaaata aaggtgagag cattcacaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2340
aaaaaaa
2347

<210> 4380
<211> 652
<212> PRT
<213> Homo sapiens

<400> 4380
Met Glu Lys Gly Glu Asp Pro Val Pro Thr Cys Leu Thr Arg Thr Gly
1 5 10 15
Leu Phe Leu Arg Phe Leu Cys Ser Arg Phe Pro Arg Gly Ala Gln Leu
20 25 30
Arg Gly Ala Leu Arg Thr Leu Ser Leu Leu Ala Ala Gln Gly Leu Trp
35 40 45
Ala Gln Thr Ser Val Leu His Arg Glu Asp Leu Glu Arg Leu Gly Val
50 55 60
Gln Glu Ser Asp Leu Arg Leu Phe Leu Asp Gly Asp Ile Leu Arg Gln
65 70 75 80
Asp Arg Val Ser Lys Gly Cys Tyr Ser Phe Ile His Leu Ser Phe Gln
85 90 95
Gln Phe Leu Thr Ala Leu Phe Tyr Thr Leu Glu Lys Glu Glu Glu Glu
100 105 110
Asp Arg Asp Gly His Thr Trp Asp Ile Gly Asp Val Gln Lys Leu Leu
115 120 125
Ser Gly Val Glu Arg Leu Arg Asn Pro Asp Leu Ile Gln Ala Gly Tyr

tgcttggcca agaaccccat tgggaataca ggggtgaagt ttctgtgtga gggcttgagg
1560
taccccgagt gtaaactgca gaccttggtg ctttggaact gcgacataac tagcgatggc
1620
tgctgcgatc tcacaaagct tctccaagaa aaatcaagcc tgttgtgttt ggatctgggg
1680
ctgaatcaca taggagttaa gggaatgaag ttcctgtgtg aggctttgag gaaaccactg
1740
tgcaacttga gatgtctgtg gttgtgggga tgttccatcc ctccgttcag ttgtgaagac
1800
gtctgctctg ccctcagctg caaccagagc ctcgtcactc tggacctggg tcagaatccc
1860
ttggggtcta gtggagtgaa gatgctgttt gaaaccttga catgttccag tggcaccctc
1920
cggacactca ggttgaaaat agatgacttt aatgatgaac tcaataagct gctggaagaa
1980
atagaagaaa aaaacccaca actgattatt gatactgaga aacatcatcc ctgggaagaa
2040
aggccttctt ctcatgactt catgatctga atccccccga gtcattcatt ctccatgaag
2100
tcatcgattt tccaggtgtg ggtgaactgc ctgtgactcc tctcctcccc cgcccctacc
2160
cctcagggat aatgagttca ttgctgggct agatgtttta gccatgattc tgcctctgtt
2220
ttatacctgc acacatcctt atctttgtta catatgaaat atctgtatca cgggtatatt
2280
gagagaaata aaggtgagag cattcacaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2340
aaaaaaa
2347

<210> 4380
<211> 652
<212> PRT
<213> Homo sapiens

<400> 4380
Met Glu Lys Gly Glu Asp Pro Val Pro Thr Cys Leu Thr Arg Thr Gly
1 5 10 15
Leu Phe Leu Arg Phe Leu Cys Ser Arg Phe Pro Arg Gly Ala Gln Leu
20 25 30
Arg Gly Ala Leu Arg Thr Leu Ser Leu Leu Ala Ala Gln Gly Leu Trp
35 40 45
Ala Gln Thr Ser Val Leu His Arg Glu Asp Leu Glu Arg Leu Gly Val
50 55 60
Gln Glu Ser Asp Leu Arg Leu Phe Leu Asp Gly Asp Ile Leu Arg Gln
65 70 75 80
Asp Arg Val Ser Lys Gly Cys Tyr Ser Phe Ile His Leu Ser Phe Gln
85 90 95
Gln Phe Leu Thr Ala Leu Phe Tyr Thr Leu Glu Lys Glu Glu Glu Glu
100 105 110
Asp Arg Asp Gly His Thr Trp Asp Ile Gly Asp Val Gln Lys Leu Leu
115 120 125
Ser Gly Val Glu Arg Leu Arg Asn Pro Asp Leu Ile Gln Ala Gly Tyr

```
    130                 135                 140
Tyr Ser Phe Gly Leu Ala Asn Glu Lys Arg Ala Lys Glu Leu Glu Ala
145                 150                 155                 160
Thr Phe Gly Cys Arg Met Ser Pro Asp Ile Lys Gln Glu Leu Leu Arg
                165                 170                 175
Cys Asp Ile Ser Cys Lys Gly Gly His Ser Thr Val Thr Asp Leu Gln
            180                 185                 190
Glu Leu Leu Gly Cys Leu Tyr Glu Ser Gln Glu Glu Glu Leu Val Lys
        195                 200                 205
Glu Val Met Ala Gln Phe Lys Glu Ile Ser Leu His Leu Asn Ala Val
    210                 215                 220
Asp Val Val Pro Ser Ser Phe Cys Val Lys His Cys Arg Asn Leu Gln
225                 230                 235                 240
Lys Met Ser Leu Gln Val Ile Lys Glu Asn Leu Pro Glu Asn Val Thr
                245                 250                 255
Ala Ser Glu Ser Asp Ala Glu Val Glu Arg Ser Gln Asp Asp Gln His
            260                 265                 270
Met Leu Pro Phe Trp Thr Asp Leu Cys Ser Ile Phe Gly Ser Asn Lys
        275                 280                 285
Asp Leu Met Gly Leu Ala Ile Asn Asp Ser Phe Leu Ser Ala Ser Leu
    290                 295                 300
Val Arg Ile Leu Cys Glu Gln Ile Ala Ser Asp Thr Cys His Leu Gln
305                 310                 315                 320
Arg Val Val Phe Lys Asn Ile Ser Pro Ala Asp Ala His Arg Asn Leu
                325                 330                 335
Xaa Pro Xaa Ala Leu Arg Gly His Lys Thr Val Thr Tyr Leu Thr Leu
            340                 345                 350
Gln Gly Asn Asp Gln Asp Asp Met Phe Pro Ala Leu Cys Glu Val Leu
        355                 360                 365
Arg His Pro Glu Cys Asn Leu Arg Tyr Leu Gly Leu Val Ser Cys Ser
    370                 375                 380
Ala Thr Thr Gln Gln Trp Ala Asp Leu Ser Leu Ala Leu Glu Val Asn
385                 390                 395                 400
Gln Ser Leu Thr Cys Val Asn Leu Ser Asp Asn Glu Leu Leu Asp Glu
                405                 410                 415
Gly Ala Lys Leu Leu Tyr Thr Thr Leu Arg His Pro Lys Cys Phe Leu
            420                 425                 430
Gln Arg Leu Ser Leu Glu Asn Cys His Leu Thr Glu Ala Asn Cys Lys
        435                 440                 445
Asp Leu Ala Ala Val Leu Val Val Ser Arg Glu Leu Thr His Leu Cys
    450                 455                 460
Leu Ala Lys Asn Pro Ile Gly Asn Thr Gly Val Lys Phe Leu Cys Glu
465                 470                 475                 480
Gly Leu Arg Tyr Pro Glu Cys Lys Leu Gln Thr Leu Val Leu Trp Asn
                485                 490                 495
Cys Asp Ile Thr Ser Asp Gly Cys Cys Asp Leu Thr Lys Leu Leu Gln
            500                 505                 510
Glu Lys Ser Ser Leu Leu Cys Leu Asp Leu Gly Leu Asn His Ile Gly
        515                 520                 525
Val Lys Gly Met Lys Phe Leu Cys Glu Ala Leu Arg Lys Pro Leu Cys
    530                 535                 540
Asn Leu Arg Cys Leu Trp Leu Trp Gly Cys Ser Ile Pro Pro Phe Ser
545                 550                 555                 560
Cys Glu Asp Val Cys Ser Ala Leu Ser Cys Asn Gln Ser Leu Val Thr
```

```
                565                 570                 575
Leu Asp Leu Gly Gln Asn Pro Leu Gly Ser Ser Gly Val Lys Met Leu
            580                 585                 590
Phe Glu Thr Leu Thr Cys Ser Ser Gly Thr Leu Arg Thr Leu Arg Leu
        595                 600                 605
Lys Ile Asp Asp Phe Asn Asp Glu Leu Asn Lys Leu Leu Glu Glu Ile
    610                 615                 620
Glu Glu Lys Asn Pro Gln Leu Ile Ile Asp Thr Glu Lys His His Pro
625                 630                 635                 640
Trp Glu Glu Arg Pro Ser Ser His Asp Phe Met Ile
                645                 650
```

<210> 4381
<211> 1638
<212> DNA
<213> Homo sapiens

<400> 4381
nnagagcccg gggcgagtgg gcctctgctc gtgggtggtt ctcgtggagg tcagctcccg
60
cgtgtctccg ctcgacaggg tgcttgggca gagcccatcg ggtaggcgcg ggccatggcg
120
cagtacaagg gcaccatgcg cgaggcaggc cgtgccatgc acctcctcaa gaagcgcgaa
180
aggcagcggg agcagatgga ggtgctgaag cagcgcatcg ccgaggagac catcctcaag
240
tcgcaggtgg acaagaggtt ctcggcgcat tacgacgccg tggaggccga gctgaagtcc
300
agcgcggtgg gcctggtgac cctgaacgac atgaaggccc ggcaggaggc cctggtcagg
360
gagcgcgagc ggcagctggc caagcgccag cacctggagg agcagcggct gcagcaggag
420
cggcagcggg agcaggagca gcggcgcgag cgcaagcgta agatctcctg cctgtccttt
480
gcactagacg acctcgatga ccaggccgac gcggccgagg ccaggcgcgc cggaaacctg
540
ggcaagaacc ccgacgtgga caccagcttc ctgccagacc gcgaccgcga ggaggaggag
600
aaccggctcc gagaggagct gcgccaagag tgggaggcgc agcgcgagaa agtgaaggac
660
gaggagatgg aggtcacctt cagctactgg gacggctcgg gccaccggcg cacggtgcgg
720
gtgcgcaagg gcaacacggt gcagcagttc ctgaagaagg cgctgcaggg gctgcgcaag
780
gacttcctgg agctgcgctc cgccggcgtg gagcagctca tgttcatcaa ggaggacctc
840
atcctgccgc actaccacac cttctacgac ttcatcatcg ccagggcgag gggcaagagc
900
gggccgctct tcagcttcga tgtgcacgat gacgtgcgcc tgctcagcga cgccaccatg
960
gagaaggacg agtcgcacgc gggcaaggtg gtgctgcgca gctggtacga gaagaacaag
1020
cacatcttcc ccgccagccg ctgggaggcc tatgaccccg agaagaagtg ggacaagtac
1080